DOI:
10.1093/medlaw/fwz016

Document Version
Peer reviewed version

This paper has been accepted for publication by the Medical Law Review. This is not the final version of the paper. For the final version, and for citation purposes, please refer to the publisher's website.

BIG DATA GOVERNANCE NEEDS MORE COLLECTIVE RESPONSIBILITY: THE ROLE OF HARM MITIGATION IN THE GOVERNANCE OF DATA USE IN MEDICINE AND BEYOND

Aisling McMahon, Alena Buyx, Barbara Prainsack*

ABSTRACT

Harms arising from digital data use in the big data context are often systemic and cannot always be captured by linear cause and effect. Individual data subjects and third parties can bear the main downstream costs arising from increasingly complex forms of data uses – without being able to trace the exact data flows. Because current regulatory frameworks do not adequately address this situation, we propose a move towards harm mitigation tools to complement existing legal remedies. In this article, we make a normative and practical case for why individuals should be offered support in such contexts and how harm mitigation tools can achieve this. We put forward the idea of 'Harm Mitigation Bodies' (HMBs) which people could turn to who feel they were harmed by data use but do not qualify for legal remedies, or that existing legal remedies do not address their specific circumstances. HMBs would help to obtain a better understanding of the nature, severity, and frequency of harms occurring from both legal and illegal data use, and they could also provide financial support in some cases. We set out the role and form of these HMBs for the first time in this article.

Keywords: Big Data; Biomedical data; Collective responsibility and oversight; Data Governance; GDPR; Harm mitigation

Mustafa's case

Mustafa loves good coffee. In his free time, he often browses high-end coffee machines that he cannot currently afford but is saving for. One day, travelling to a friend's wedding abroad, he gets to sit next to another friend on the plane. When Mustafa complains about how much he paid for his ticket it turns out that his friend paid less than half of what he paid. Mustafa googles possible reasons for this and concludes that it must be related to his browsing expensive coffee machines and equipment. He is very angry about this and complains to the airline, who send him a lukewarm apology that refers to their pricing being based on demand. Mustafa feels that this is unfair but lets it go because pursuing this would make him lose time and money.

Paula's case[1]

After years of trying to conceive, Paula is pregnant. Five months into the pregnancy the unspeakable happens and she loses the baby. Paula and her partner are heartbroken. For months after the end of her pregnancy Paula keeps receiving advertisements from shops specialised on maternity and infant products and services congratulating her on the 'milestones' of her supposed baby. This is an immensely aggravating and distressing experience for Paula and her partner. Paula's partner calls up the companies that send these advertisements, demanding them to erase their names from their database. He also demands to hear where they got Paula's contact details in the first place, but he does not receive any answers. Paula suspects that one of her doctors passed on her details to retailers, but she cannot prove it.

*The authors would like to thank TT Arvind, Rebekah Farrell, Carrie Friese, Hanna Kienzler, Thomas King (UK Royal Statistical Society) and Graeme Laurie for their very helpful comments on earlier drafts. They are also grateful to the comments of anonymous reviewers. The usual disclaimer applies.

[1] This vignette is inspired by Mary Ebeling's important book: Mary F. Ebeling, *Healthcare and Big Data* (New York: Palgrave Macmillan US, 2016).

INTRODUCTION

These brief vignettes demonstrate the potential for individuals to be harmed by data use in the big data era particularly, given the pervasive nature of such data. The 'pervasiveness'[2] of data means that in the context of large, digital, connected or connectable, and often fast-changing datasets, harms are often systemic and cannot be captured by linear cause and effect. Downstream harms arising from digital data use can fall outside of the remit of traditional legal remedies because they do not fit the traditional chain of causality which legal actions typically require – such as under: the tort of misuse of private information in the UK context,[3] or for some actions under the new European Union's General Data Protection Regulation (GDPR)[4], or for the right to private and family life (Art 8 European Convention on Human Rights). Specifically, people who experience harms that they can plausibly assume stems from data use may not have access to legal remedies for two reasons: First, because the action that led to the harm was lawful – such as in the case of Mustafa, who was affected by personalised pricing practices that many people may consider unfair,[5] but that are not illegal. Second, even if it is apparent that the action that caused the harm must have been unlawful (such as in Paula's case, where somebody must have infringed data protection laws and passed on her information to retailers) it may be impossible to prove that a specific instance of data use caused a specific

[2] See <https://pervade.umd.edu/about/data-ethics-regulators/> accessed 8th April 2019 and Jacob Metcalf, '"The study has been approved by the IRB": Gayface AI, research hype and the pervasive data ethics gap' (2017) Medium (30 November) < https://medium.com/pervade-team/the-study-has-been-approved-by-the-irb-gayface-ai-research-hype-and-the-pervasive-data-ethics-ed76171b882c> accessed 8th April 2019.

[3] This tort was recognised by the UK courts in *Vidal-Hall v Google* [2014] EWHC 13 (QB) which allowed data subjects to bring a claim against data controllers for compensation in cases of distress caused by data use. In this case the claimants' online browsing activities was tracked by Google and used to profile the claimants and then direct targeted advertisements to them. The claimants complained of the distress they suffered based on Google's use of these characteristics without their consent/knowledge to send targeted advertisements, and the risk that such personal characteristics could have come to the knowledge of third parties who used/saw their devices.

[4] Art 82(1) GDPR 2016/679 provides that: "Any person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered". The GDPR became directly applicable on 25th May 2018.

[5] See discussion in: Timothy J. Richards, Jura Liaukonyte, Nadia A.Streletskayab, 'Personalized pricing and price fairness' (2016) 44 *International Journal of Industrial Organization* 138-153.

harm and therefore the chain of causation required for legal remedies may be lacking.[6] This may be because the harm was caused not by a single specific use of that individual's data, but by the use of data from different sources in combination, or where harm to a specific individual resulted from the use of another individual's data, without the former being at all aware of this usage. Digital data are also multiple in the sense that they can be in more than one place at once, and that they can be interlinked in various ways,[7] which makes it extremely difficult to trace the movements of data in specific cases. These circumstances can leave individual data subjects, such as Mustafa and Paula, and third parties unable to prove causal links between data use(s) and the harm suffered, leaving them to bear the downstream costs in the form of harms arising from increasingly complex forms of data uses – without clear legal remedies.

At the same time, downstream costs on data controllers, who often derive significant commercial benefits from such uses, are limited. This amounts to a significant imbalance of benefits and harms, and, more broadly speaking, of power between data users and data subjects or third parties.

In view of this situation, a recalibration of data governance is necessary. As part of a larger programme of thinking about data-governance,[8] this article makes the case for harm mitigation

[6] The principles of accountability and responsibility under the GDPR do not displace this point. This is because for compensation, the controller/processor is exempt only if they can prove that the damage is not linked to their actions. Damage assessment thus depends on causation, as do other aspects.

[7] Barbara Prainsack, 'Data donation: How to resist the iLeviathan' in Jenny Krutzinna and Luciano Floridi (eds) *The Ethics of Medical Data Donation* (Springer, 2019) at 9-22.

[8] The programme is outlined in Barbara Prainsack & Alena Buyx, 'A solidarity-based approach to the governance of research biobanks' (2013) 21(1) *Medical Law Review* 71-91; Barbara Prainsack and Alena Buyx, *Solidarity in Biomedicine and Beyond* (Cambridge University Press, 2017); Barbara Prainsack and Alena Buyx, 'Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic' (2016) 37 *Theoretical Medicine and Bioethics* 489-501; Barbara Prainsack, 'Research for personalised medicine: Time for solidarity' (2017). 36(1) *Medicine and Law* 87-98; Gesine Richter, Christoph Borzikowski, Wolfgang Lieb, Stefan Schreiber, Michael Krawczak, Alena Buyx, 'Patient views on research use of clinical data without consent: Legal, but also acceptable?' 25.01.2019 *European Journal of Human Genetics*; Amelia Fiske, Alena Buyx, Barbara Prainsack, 'Health Information Counselors: A New Profession for the Age of Big Data?' (2019) 94(1) *Academic Medicine* 37-41; Gesine Richter, Michael Krawczak, Wolfgang Lieb, Lena Wolff, Stefan Schreiber, Alena Buyx , 'Broad consent for healthcare-embedded biobanking: understanding and reasons to donate in a large patient sample' 20 (1) *Genetics in Medicine* 76-82.

tools to complement existing legal frameworks. In doing so, we put forward a normative and practical rationale for why individuals should be offered support in such contexts, and why systems for monitoring such 'harms' should be established. We envisage these functions being conducted by a 'Harm Mitigation Body' (HMB), which could also establish financial support mechanisms. While two of the authors have sketched the general idea of such bodies in previous work,[9] we set out the role and form of these HMBs for the first time in this article.

At the outset, we also acknowledge that the adoption of the GDPR takes positive steps to address some of the issues caused by big data practices. For example, the GDPR's principle of accountability of data controllers moves the onus of proof onto data controllers thereby reducing the need for data subjects to demonstrate causation in many contexts. However, as will be demonstrated, gaps remain and the GDPR's remit is still too narrow to provide effective harm mitigation for all data subjects. To address this, the proposed HMB framework could provide support for harms caused through data use: (a) in cases where data use did not breach the GDPR; and (b) in countries where the GDPR is not applicable. Moreover, HMBs would not focus on policing (and placing fines on) data controllers but on providing support to data subjects. They thus complement, rather than duplicate or compete with, the institutions and instruments of the GDPR.

In making these arguments, the article is structured as follows: Part I sets out the value of 'big data' in today's world and the challenges it poses for governance. It puts forward a normative and practical case for a renewed focus on the need for harm mitigation, and our novel tool to do so, the HMB. Following this, Part II describes the main functions and legal adaptations of

[9] Barbara Prainsack & Alena Buyx, 'A solidarity-based approach to the governance of research biobanks' (2013) 21(1) *Medical Law Review* 71-91; Barbara Prainsack and Alena Buyx, *Solidarity in Biomedicine and Beyond* (Cambridge University Press, 2017).

such HMBs, setting out an operational overview illustrating how these bodies would work within the national data protection context, and their structural composition etc.

While the arguments were initially developed in the context of the governance of data for biomedical research and practice, the harm mitigation framework spelled out here is not limited to the medical domain, but, in principle, applicable to any instance of data use.

I. BACKGROUND: BIG DATA AND THE NEED FOR NEW GOVERNANCE FRAMEWORKS

A. The context: Big Data in biomedicine

'Big data'[10] are a key resource for medical research and increasingly also medical practice in the digital era. Owing to advances in information technologies it has become easier to link data from multiple sources and datasets in recent years,[11] with many knock-on benefits for medical research. In the definition of the Garter IT glossary, which the UK Information Commissioner's Office refers to, big data are "…high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[12] Big data practices and epistemologies are hoped to help improve the

[10] Defined by the EU Commission as: "large amounts of different types of data produced from various types of sources, such as people, machines or sensors. This data could be climate information, satellite imagery, digital pictures and videos, transition records or GPS signals. Big Data may involve personal data: that is, any information relating to an individual, and can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address", see EU Commission, 'The EU Data Protection Reform and Big Data' (January 2016).

[11] Barbara Prainsack & Alena Buyx, Solidarity in biomedicine and beyond (Cambridge: Cambridge University Press, 2017), 11; Henry Pearce, 'A systems approach to data protection law and policy in the world of big data?' (2016) 22(4) Computer and Telecommunications Law Review 90-93.

[12] Gartner IT glossary, 'Big data' as cited by the UK Information Commissioner Office, who have conceded that this definition based on the three v's (volume, velocity and variety) has been described as tired through over-use and can be problematic as multiple forms of data do not share the same traits. They supplement it by noting that it "is useful to regard it as data which, due to several varying characteristics, is difficult to analyse using traditional data analysis methods" Information Commissioner's Office, 'Big data, artificial intelligence, machine learning and data protection' (September 2017) Version 2.2 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.> accessed 8th April 2019, 6.

efficiency and effectiveness in fostering healthy habits/practices,[13] to enable more precise prediction and more successful prevention of disease, and aid the development of medical interventions,[14] both from a commercial point of view as well as from a public health perspective. Artificial intelligence[15] and machine learning[16] techniques are useful to accelerate these benefits and fully "unlock the value of big data".[17] Personal data– understood, in line with Art. 4 of the GDPR as information relating to an identified or identifiable natural person, play important roles in this endeavour.

At the same time, traditional distinctions drawn by data governance frameworks, such as the distinctions between identified (or identifiable) and anonymous data,[18] and between sensitive

---

Some have also discussed definitions based on four and five v's adding value and veracity, see: Jonathan Shaw, 'Why "Big Data" Is a Big Deal: Information Science Promises to Change the World', Harv. Mag. (Mar.-Apr. 2014) <http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal> accessed 21 May 2018; Steve Lohr, 'The Age of Big Data', NY Times (Feb. 11, 2012) <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html > accessed 21 May 2018; Svetlana Sicular, 'Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s' FORBES (Mar. 27, 2013) < https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/ > accessed 8th April 2019. See also Big Data, GARTNER < https://www.gartner.com/it-glossary/big-data / > accessed 8th April 2019; Chris Forsyth, 'For Big Data Analytics There's No Such Thing as Too Big: The Compelling Economics and Technology of Big Data Computing', 4SYTHCOMM.COM (Mar. 2012), <https://www.cisco.com/c/dam/en/us/solutions/data-center-virtualization/big_data_wp.pdf > .as cited in TZ Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2018) 47 Seton Hall Review 995, 999.

[13] Fabrico F Costa 'Big Data in biomedicine' (2014) 19(4) Drug Discovery Today 433–440.

[14] Charles Auffray et al, 'Making sense of big data in health research: Towards an EU action plan' (2016) 71(8) Genomic Medicine 1-13, 3.

[15] Defined as "…the analysis of data to model some aspect of the world. Inferences from these models are then used to predict and anticipate possible future events" in Government Office for Science, Artificial intelligence: opportunities and implications for the future of decision making, (9 November 2016) as cited by Information Commissioner's Office, Big data, artificial intelligence, machine learning and data protection (September 2017) Version 2.2., 6.

[16] Defined as "the set of techniques and tools that allow computers to 'think' by creating mathematical algorithms based on accumulated data" in Deb Landau, Artificial Intelligence and Machine Learning: How Computers Learn. iQ (17 August 2016) < as cited in the Information Commissioner's Office, Big data, artificial intelligence, machine learning and data protection (September 2017) Version 2.2., 7.

[17] Information Commissioner's Office, 'Big data, artificial intelligence, machine learning and data protection' (September 2017) Version 2.2, 8

[18] On the difficulties in anonymisation in big data era, see: Barbara Prainsack & Alena Buyx, Solidarity in biomedicine and beyond. (Cambridge: Cambridge University Press, 2017) 11. See also: Graeme Laurie & Leslie Stevens, 'Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom' (2016) 44(3) Journal of Law and Society 360-392, 368 citing the following: P. Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Rev. 1701; A. Narayanan & V. Shmatikov, `De-Anonymizing Social Networks' (2009) 30th IEEE Symposium on Security & Privacy, at <https://www.cs.utexas.edu/~shmat/shmat_oak09.pdf>; P. Schwartz & D. Solove, `The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 New York University Law Rev. 1814; M. Gymrek et al., `Identifying Personal Genomes by Surname Inference' (2013) 339 Science 321.

and non-sensitive data[19] are increasingly difficult to operationalise in the big-data era given the increased sharing, copying, and linking of data and datasets, and because of the aforementioned multiplicity of digital data – the fact that they can be in more than one place at the same time. If combined with sufficient additional data and information, virtually any data point is identifiable. Moreover, data representing even the most innocuous kind of information could be used in conjunction with other data to reveal sensitive information thereby increasing the risk of harming people.[20] It has been argued that, in the digital era, any data should be regarded as potentially identifiable, health-related, and sensitive.[21] In addition, big data allows greater emphasis on 'insights obtained at the aggregate level to be used to make probabilistic "predictions"' about individuals and groups,[22] leading to challenging questions about who gets to make – and further use – such predictions, for what purposes, and under which safeguards.

B. The Problem: Challenges for the governance of human data left unaddressed by the GDPR

The regulatory response to challenges posed by big data has tended towards the fossilisation, imitation or mimesis[23] of traditional concepts and instruments such as privacy, informed consent and risk management. Broadly speaking, it has so far sought to address challenges from big data use by trying to minimise and manage risks of data use as much as possible, and, at the same time, by attempting to increase individual control over how specific data types are used.

[19] For a full discussion of sensitive data and difficulties of definition Stephanie OM Dyke, Edward S Dove, Bartha M Knoppers, 'Sharing Health Related Data: a privacy test?' (2016) NPJ Genomic Medicine; Jacob Metcalf, Emily F Keller and Danah Boyd, 'Perspectives on Big Data, Ethics and Society' The Council for Big Data, Ethics and Society (23rd May 2016) <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/> accessed 8th April 2019.

[20] For the varied types and severity of harms which can be caused by big data see Joanna Reddan and Jessica Brand, 'Data Harm Record' <https://datajusticelab.org/data-harm-record/> accessed 8th April 2019.

[21] Barbara Prainsack & Alena Buyx, Solidarity in biomedicine and beyond. (Cambridge: Cambridge University Press, 2017), 11. For an interesting discussion of these questions, see Miranda Mourby et al, 'Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK' (2018) 34(2) Computer Law and Security Review 222-233.

[22] Barbara Prainsack and Alena Buyx, Solidarity in Biomedicine and Beyond (Cambridge University Press, 2017) 97.

[23] On the idea of mimesis see: G Laurie, 'Liminality and the Limits of Law in Health Research Regulation: What are we Missing in the Spaces In-between?' (2017) Medical Law Review, 47-72, 58.

This has led to what some have termed 'privacy protectionism'[24] or a 'whiplash effect'[25] where 'overly restrictive measures (especially legislation and policies) are proposed in reaction to perceived harms, which overreact in order to re-establish the primacy of threatened values, such as privacy'.[26]At the societal level, an overt focus on risk management can increase burdens for data processing which has the potential to limit data uses, thereby hampering the potential benefits of big data - while such measures may not necessarily protect individuals either.

In addition, at an individual level, focusing on risk management and relying on traditional concepts such as informed consent is problematic because it could give false assurances. Even with the highest safeguards and best intentions in place, no data usage in the digital age will ever be entirely risk-free;[27] nor can there be a guarantee against predictive use of individual and collective health (and other) information in ways that can harm people. The emphasis put on increased individual control and risk minimisation in many current health data governance systems, as useful as this approach has been to date, is likely to also engender problematic expectations for data-subjects in the new 'big data' context, who may feel falsely assured that they have meaningful control over how their data are used.[28] A more explicit focus on harms and harm mitigation, would be in line with the principle of veracity in the governance of personal data in the digital era.[29]

---

[24] J. Allen, C.D.J. Holman, E.M. Meslin, F. Stanley, 'Privacy protectionism and health information: any redress for harms to health?' (2013) 21(2) J. Law Med. 473–485 as cited in K.H. Jones, G Laurie, L Stevens, C Dobbs, DV Ford, N Lea, 'The other side of the coin: Harm due to the non-use of health-related data' (2017) 97 International Journal of Medical Informatics 43–51, 47.

[25] Brent Daniel Mittelstadt and Luciano Floridi, 'Introduction' in *The Ethics of Biomedical Big Data* (Springer International Publishing, 2016), 1.

[26] Ibid, 1.

[27] See Joanna Reddan and Jessica Brand, 'Data Harm Record' https://datajusticelab.org/data-harm-record/ accessed 8th April 2019; See also work of PERVADE group on pervasive nature of big data <https://pervade.umd.edu/> accessed 8th April 2019.

[28]Barbara Prainsack & Alena Buyx, Solidarity in biomedicine and beyond. (Cambridge: Cambridge University Press, 2017) 115.

[29] Jeantine E Lunshof et al, 'From genetic privacy to open consent' (2008) 9(5) *Nature Reviews Genetics* 406-411.

More effective harm mitigation mechanisms are particularly important and timely also because big data usage can lead to harms that have so far not been recognised, such as those described in Paula's and Mustafa's case in the beginning of the article. Predictive analytics could also lead to poor consumer ratings which in turn result in the denial of mortgages; predictive health profiles could make finding employment more difficult or raise insurance premiums, etc.[30] Studies have also shown the risks of 'credit worthiness by association'[31] such as where an individual's credit card limit was reduced due to predictions based on repayment histories of other people who shopped in the same stores as he had.[32] Importantly, harms can occur also when data processing is lawful, such as in Mustafa's case, which is a situation that existing legal remedies do not address.[33]

Legal systems have tried to adapt to some of the challenges posed by big data. At a European level, the GDPR under Recital 71 protects against the processing of data which causes discriminatory effects on people on the basis of "racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect." However, there are clear limitations to such measures given the way big data works. The Information Commissioner Office (UK)

---

[30] See generally: Joanna Redden, 'Six ways (and counting) that big data systems are harming society' (7th December, 2017) The Conversation <https://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660?utm_content=buffer0c4fa&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer> accessed 8th April 2019, which cites: Federal Trade Commission, 'Big Data: A Tool for Inclusion or Exclusion?' Understanding the Issues (January, 2016) <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> accessed 8th April 2019 ; Solon Barocas & Andrew D. Selbst, Big Data's Disparate Impact, (2016) 104 California Law Review 671; Danielle Keats Citron & Frank A Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1.

[30] Joanna Redden, 'Six ways (and counting) that big data systems are harming society' (7th December, 2017) The Conversation <https://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660?utm_content=buffer0c4fa&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer> accessed 8th April 2019.

[31] Ibid.

[32] Mikella Hurley & Julius Adebayo, Julius, 'Credit Scoring in the era of Big Data' (2016) 18(1) Yale Journal of Law and Technology 148 http://digitalcommons.law.yale.edu/yjolt/vol18/iss1/5 accessed 8th April 2019..

[33] For a running record of data harm examples see https://datajusticelab.org/data-harm-record/.

has stated that big data analysts "will need to find ways to build discrimination detection into their machine learning systems to prevent such decisions being made in the first place"[34] to comply with issues of accountability under GDPR. Given the way the technology works, not all potential discriminatory effects will be reasonably foreseeable to allow pre-emptive action.[35]

The GDPR also gives natural persons the right not to be subjected to decisions based on profiling under certain circumstances.[36] There are, however, numerous exemptions, and limitations to this, discussed in further detail below.[37] It is also noteworthy that the respective Articles of the GDPR (4(4), 9, 22; see also Recitals 71-2) do not protect individuals from their personal information being used for automated purposes *per se*; rather it protects them merely from being subjected to decisions based *solely* on such automated purposes, e.g. profiling, without human intervention.[38] The result is that the prohibition can easily be bypassed by involving humans at some stage in a largely automated process of decision making, for example by signing off results suggested by an algorithm.[39]

[34] Information Commissioner's Office, 'Big data, artificial intelligence, machine learning and data protection' (September 2017) Version 2.2 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.> accessed 8th April 2019, para 116.

[35] See discussion of risks of discrimination in: Rhoen and Feng, 'Why the 'Computer says no': illustrating big data's discrimination risk through complex systems science' (2018) 8(2)(1) *International Data Privacy Law* 140–159,

[36] Profiling, here, is defined in Art 4(4) of the GDPR as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"

[37] See also Michael Vaele & Lillian Edwards, 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling' (2018) 34(2) Computer Law & Security Review 398-404.

[38] Rita Heimes, 'Top 10 operational impacts of the GDPR: Part 5 – Profiling' The Privacy Advisor (20th January 2016) <https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-5-profiling/> accessed 8th April 2019.

[39] Tal Z Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2018) 47 Seton Hall Review 995, 1016-1017. This can arguably be inferred from the discussion of automated decision-making in: Art 29 Working Party, Guidelines on Automated Individual decision-making and profiling for the purposes of Regulation 2016/679 adopted on 3 February 2017 WP251.01 available at https://iapp.org/media/pdf/resource_center/W29-auto-decision_profiling_02-2018.pdf These guidelines were endorsed by the European Data Protection Board (EDPB) in May 2018 https://www.dataprotection.ie/en/European-Data-Protection-Board

Furthermore, under the GDPR, data can only be processed in line with the purpose it is collected for (purpose limitation),[40] but subsequent processing is permissible provided it is compatible with this purpose. One such compatible purpose is processing for statistical purposes, however uncertainty remains around how this exception will apply in the big data context.[41] To be deemed a statistical purpose the results should not be used "in support of measures or decisions regarding any particular natural person."[42] However, in practice it may be impossible to prevent data users to use findings from data analytics that identify an association (or even a causal connection[43]) between two parameters in such a way that they apply to specific individuals.

Either as part of implementing the GDPR into their national legal systems, or additionally, many countries currently attempt to introduce legislation curbing or at least limiting the use of predictive analytics in health care systems.[44] Although these solutions are a welcome step, they are not sufficient to address the problem described above for four reasons: First, potential harms caused by the use of data in this way often fall outside the remit of governance frameworks using traditional legal concepts.[45] As demonstrated by Metcalf and other colleagues' work on the pervasive nature of big data, harms are often systemic in nature and may have effects on multiple individuals downstream.[46] For example, a person who is harmed by a predictive

[40] Art 5(1)(b) GDPR and Art 6(4) GDPR.
[41] For a discussion of the likely difficulties this will pose for big data, see Zarsky, note 40, 1008; Antoinette Rouvroy, ""Of Data and Men": Fundamental Rights and Freedoms in a World of Big Data", *Council of Europe Directorate General of Human Rights and Rule of Law* at 11 (Jan. 11, 2016).
[42] Recital 162, GDPR. See discussion in Zarsky, 1008; Antoinette Rouvroy, ""Of Data and Men": Fundamental Rights and Freedoms in a World of Big Data", *Council of Europe Directorate General of Human Rights and Rule of Law* at 11 (Jan. 11, 2016), https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020 at 26; for an alternative view see Viktor Mayer-Schönberger & Yann Padova, Regime Change? Enabling Big Data through Europe's New Data Protection Regulation, (2016) 17 Colum. Sci. & Tech Law Rev.315 at 329
[43] Hernán, Miguel A., John Hsu, and Brian Healy. "Data science is science's second chance to get causal inference right: A classification of data science tasks." (2018) *arXiv preprint arXiv:1804.10846*.
[44] For example, see adapted German Data Protection Law.
[45] Barbara Prainsack & Alena Buyx, Solidarity in biomedicine and beyond. (Cambridge: Cambridge University Press, 2017), 98. See also: Mark Taylor, Genetic data and the law: a critical perspective on privacy protections (Cambridge: Cambridge University Press, 2013)
[46] See https://datajusticelab.org/data-harm-record/ ; See also work of PERVADE group on pervasive nature of big data - https://pervade.umd.edu/,

analytics system that makes probabilistic inferences regarding an undesirable trait on the basis of generic information about her – such as the postcode she lives in – does not have access to legal remedies if the data that was used to make these inferences is not her own personal data. Second, even when it is clear that a particular harm must have resulted from a breach of the law – such as in the case of Paula described in the beginning of the article – individuals might not be able to prove the relevant causal link required for traditional tort based remedies to apply, or for the purposes of establishing a right to compensation under the GDPR. This is because multiple digital copies of datasets may exist whose movements cannot be traced, or because the pervasive nature of digital data may make it easier for data controllers or processers to prove they are not responsible for the event. Third, according to the GDPR, even when data controllers fail to take sufficient steps to avoid discrimination against people and, as a result, they can be held accountable for breach of the GDPR, existing frameworks impose penalties for mis- and abuse of data that are often too low to deter bad practice on the side of data controllers[47] – and neither do they provide support for the data subjects that experienced the harm. And fourth, Art 80 GDPR provides that non-profit bodies active in the privacy context can initiate claims of infringement by data subjects, and allows for the potential for group actions – if such bodies gather multiple claims on similar issues from data subjects.[48] This measure could be used as a collective support mechanism for individuals, but shortfalls remain, namely: First, it will be left to each Member State to devise such mechanisms for non-profit actions in national jurisdictions

[47] Frank Pasquale, The Black Box Society: The Secret Algorithms that Control Money and Information. (Cambridge, MA: Harvard University Press, 2015), 91. Under the GDPR, organizations in breach of the GDPR can be fined up to 4% of annual global turnover or €20 Million (whichever is greater). However, whether this will be sufficient to deter mis/abuse remains to be seen.

[48] See also Recital 142, GDPR. For a discussion see: Kellie O'Flynn, 'Has the GDPR Opened the Door to Class Actions in Ireland?' (27 August 2018) <https://www.williamfry.com/newsandinsights/news-article/2018/08/27/has-the-gdpr-opened-the-door-to-class-actions-in-ireland > accessed 8th April 2019. The first such complaints under Art 80 GDPR were lodged by Max Schrems group 'None of Your Business' (NOYB) in May 2018. See https://noyb.eu/4complaints/ ; The actions were against Google, Facebook, WhatsApp and Instagram. The action against Google has led to a decision in January 2019 by CNIL (the French Data Protection Commission) fining Google 50million Euros; see https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc

which could lead to variance of approaches; Second, such bodies would also still only be able to take actions based on the infringement of the GDPR.

All these aspects underscore the need for a new body, namely an HMB. Such a body would complement regulation to police data controllers by focusing on supporting data subjects who were harmed by data use, irrespective of who caused the harm, of whether or not they are able to prove a causal connection, and irrespective of whether the action or omission that led to the harm was illegal.

C. Why better harm mitigation is needed in the digital era

Alongside other colleagues we thus argue that the current approach to data governance needs to change.[49] In previous publications two of the authors of this article (Barbara Prainsack and Alena Buyx),[50] have argued that new approaches to big data governance should be based on the concept of solidarity,[51] and that they should include three main pillars: (a) greater emphasis on whether or not specific instances of data use are in the public interest, (b) the strengthening of harm mitigation instruments, and (c) new legal mechanisms to ensure that significant parts of financial profits created on the basis of data use go into the public purse (e.g. via a corporate data use tax). In this article, we will focus on pillar (b) harm mitigation. As noted, the account

[49]See also: Frank Pasquale, The Black Box Society: The Secret Algorithms that Control Money and Information. (Cambridge, MA: Harvard University Press, 2015); Viktor Mayer-Schonberger & Kenneth Cukier, Big Data: A Revolution that will transform how we live, work and think (Houghton Mifflin Harcourt, New York, 2013); Dana Boyd & Kate Crawford, 'Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon' (2012) 15(2) Information, Communication and Society 662-679; N. Kshetri, 'The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns' (2014) 1 Big Data & Society; S. Barocas & H. Nissenbaum, 'Big Data's End Run Around Anonymity and Consent' in J Lane et al (eds), Privacy, Big Data, and the Public Good: Frameworks for Engagement, (Cambridge University Press, 2014); Paul Ohm, 'Changing the Rules: General Principles for Data Use and Analysis' in J Lane et al (eds), Privacy, Big Data, and the Public Good: Frameworks for Engagement (Cambridge University Press, 2014).

[50]Barbara Prainsack and Alena Buyx, 'A solidarity-based approach to the governance of research biobanks' (2013) 21(1) *Medical Law Review* 71-91; Barbara Prainsack and Alena Buyx, *Solidarity in Biomedicine and Beyond* (Cambridge University Press, 2017)

[51] For a definition, see Prainsack, Barbara, and Buyx, Alena, 'Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic' 37 Theoretical Medicine and Bioethics 489-501 at 493; see also: Barbara Prainsack and Alena Buyx, *Solidarity in Biomedicine and Beyond* (Cambridge University Press, 2017).

of harm mitigation that we develop in the following sections can be adapted to, and support, the governance of any data use, not limited to health, and within any kind of regulatory and governance landscape. In other words, although the idea of harm mitigation bodies (HMBs) and their accompanying financial support mechanisms was conceived in connection with solidarity-based governance,[52] enhancing harm mitigation is an essential task in the era of digital data, irrespective of whether other instruments of solidarity-based governance are employed or not.

We argue that there are important *normative and practical* rationales for establishing HMBs. At the normative level, harm mitigation forms a key step that governance frameworks need to shift towards, strengthening not only individual but also collective control and responsibility in the context of data use. We take it as a given, in the interests of social justice and fairness, that significant inequities in the distribution of benefits and burdens, as well as large power imbalances in societies, need to be remedied. Improved mechanisms and instruments of collective control and responsibility for data use are necessary to counteract the growing imbalances between those who give data and those who use them. Strengthening collective responsibility also includes improving the support of individuals and groups of people[53] who are harmed by data use. And this, in turn, leads us to emphasise the need for better harm mitigation instruments.

In our view, HMBs will help to protect people from the downstream costs – in the widest sense of the word – of personal data use in the digital era to a greater extent than is currently the case.

[52] Barbara Prainsack and Alena Buyx, 'Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic' (2016) 37 *Theoretical Medicine and Bioethics* 489-501; Barbara Prainsack, 'Research for personalised medicine: Time for solidarity' (2017) 36(1) *Medicine and Law* 87-98; Barbara Prainsack and Alena Buyx, 'A solidarity-based approach to the governance of research biobanks. (2013) 21(1) *Medical Law Review* 71-91.

[53] Having said this, we anticipate group claims to be the exception. Furthermore, depending on size of group, they may already have collective agency to a certain extent, so such issues are likely to be resolved more quickly by a data controller than cases of individual harm in order to maintain the public image/legitimacy of the data controller.

For individuals, big data entails upfront costs (the costs related to curating and using datasets, which they share with public and private corporations that provide the technologies and infrastructures for the generation of data) and downstream costs in terms of harms caused by data use. Under current regulatory frameworks, as described, these downstream costs for data use in the digital era are carried largely by individual members of the public and by publics, whereas the commercial benefits unfold mostly for such privately owned corporations. HMBs would seek to reduce the downstream costs for individuals/publics, and thus contribute to a mitigation of imbalances in power and benefits and burdens between publics and private actors. While this is not an argument against private corporations making profits from personal data use in principle, the difference in power, agency, and costs borne by corporations that use data on the one hand, and the people who contribute data on the other, is a problem in need of addressing. In some of our other work,[54] we have suggested solutions to ensure that a larger share in profits come back into the public domain (e.g. via a data tax). These monies paid by corporations benefitting from the use of people's data could and should be used to support the operation of HMBs.

There is also a *practical rationale* for HMBs: we argue that HMBs are practically mandated because no matter how hard we try to reduce risks emerging from novel ways of data use, some individuals and groups will still inevitably be harmed by these practices, either intentionally or unintentionally, and both by legal and illegal data use. As noted, existing legal remedies,[55] which only protect from the downstream costs of data use which is unlawful e.g. infringes the GDPR,[56] and (in some cases) only if those harmed can prove a causal connection between an

[54] Barbara Prainsack and Alena Buyx, 'Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic' (2016) 37 *Theoretical Medicine and Bioethics* 489-501; Barbara Prainsack, 'Research for personalised medicine: Time for solidarity.' (2017) 36(1) *Medicine and Law* 87-98.

[55] For example, the tort of misuse of private information in the United Kingdom, or data subjects' rights under chapter 3 GDPR and remedies in cases of unlawful processing article 8 GDPR.

[56] Furthermore, scenarios could potentially arise where fines were imposed for breach of GDPR processes, but individual would not necessarily obtain compensation unless infringement leads to damage.

act or omission by a specific entity and the harm incurred, are not sufficient to address this problem. Further difficulties stem from the fact that costs often arise not only for primary data subjects but for third parties, i.e., for other individuals downstream.[57] To address these issues on a practical level, we need an instrument ancillary to traditional legal mechanisms, with an explicit focus on harm mitigation. For this system to be effective, it must be easy for people to use, very low on bureaucracy, and flexible enough in its decision-making system to support people where and how they most need it.

Moreover, although it might be possible that other avenues could be developed to address the issues described, such as strict liability for data mis-use or any use which causes harm, it is well known by now that serious harms can stem from data use that does not break any laws. While the harm experienced by Mustafa in the beginning of the article could be seen as very minor, other legal practices can cause serious harms, without effective and accessible legal remedies being available to those harmed: A man whose driver's license was revoked after facial recognition technology wrongfully 'identified' his photograph as similar to another license holder and there was suspicion of identity fraud. [58] Another person who typed something into Google that the company's autocomplete function added the word 'bomb' to, who was visited by government investigators and lost his job as a result. In both cases,[59] serious harm occurred without either of the men having access to effective legal remedies to provide financial support for the harm suffered.[60]

[57] See also PERVADE project <https://pervade.umd.edu/about/data-ethics-regulators/> accessed 8th April 2019.
[58] See discussion in Luke Dormehl, 'Algorithms are great and all but they can also ruin lives' (Wired, 19 November, 2014) available at https://www.wired.com/2014/11/algorithms-great-can-also-ruin-lives/ (<accessed 8th April 2019>
[59] See discussion of these and other similar cases in L Dormehl, The Formula: How Algorithms Solve all our Problems ... and Create More (Random House, 2014).
[60] The victim of the facial recognition mistake was told by authorities that he had the right to request a hearing but the onus was upon him to prove his identity if he wanted his licence restored. The hearing was held 11 days after the licence suspension where he submitted relevant documentation and his licence was reinstated. Nonetheless, the fact the licence was suspended in the first place due to algorithmic misidentification caused

We argue that a harm mitigation model is the most appropriate in the context described. This is because it could operate flexibly and would offer an important balance of supporting individuals who may suffer harms and seeking to limit these harms, regardless of whether these harms stem from legal or illegal data uses, and irrespective of whether the harmed individuals can prove a causal connection between a specific act or omission and the harm incurred. Furthermore, the more informal nature of HMBs, as laid out in the following, may help to address the problem that it is often the most socio-economically deprived groups that are most likely to be harmed by data uses. These groups regularly have fewer resources (both in terms of financial resources and access to legal advice) to avail themselves of traditional legal remedies and mount legal challenges against particular data uses, thus potentially further worsening the aforementioned imbalances. Finally, the HMB model would provide an avenue for individuals (and groups) to feed issues of big data usage back to regulators, thus forming an important element of reflexive governance.[61] This is vital given the evolving nature of data uses, and how, relatively, little we know at this point about the nature and severity of harms that stem from these uses, as well as their prevalence and distribution within and across populations.

II: HARM MITIGATION BODIES: FUNCTIONS AND LEGAL ADAPTATIONS

This section introduces HMBs and expands on their main functions in reference to the GDPR to highlight the complementary nature HMBs could play in addressing gaps in existing

unnecessary harm for him. His subsequent claim for damages based on the incident against the Register for Motor Vehicles was rejected. See Gass v. Registrar of Motor Vehicles 12-P-205 (Mass. App. Ct. Jan. 7, 2013). The fired employee who wanted to build a radio-controlled airplane and was accused of trying to build a bomb subsequently commenced expensive litigation to try to claim compensation for the job loss suffered.

[61] For a discussion of reflexive governance see: G Laurie, 'Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law' (2011) 130(3) Human Genetics 347; S Harmon, G Laurie, G Haddow, 'Governing risk, engaging publics and engendering trust: New horizons for law and social science? '(2013) 40(1) Science and Public Policy 25.

frameworks. We expect these ideas to be developed further over time by and in response to the community within which HMBs would operate and apply.

A. Introducing Harm Mitigation Bodies

We envisage HMBs as instruments that specifically address harms to individuals that are *plausibly* connected to data use. HMBs have two primary functions, namely: (1) to provide financial support to individuals who can plausibly make a case that they suffered significant and undue harm by data use (without needing to prove wrongdoing or direct legal causation due to an acknowledgment that this is increasingly not possible in the era of digital data); and (2) to monitor harms reported as being caused by big data practices reported to and within HMBs. This information can then be fed back to data controllers and to public agencies and inform how the operation of systems of data governance could be improved.

HMBs would be established at national levels, for example as independent statutory arms of national data protection bodies.[62] They would have oversight for data uses pertaining to all data controllers resident (in the case of individuals) or established (in the case of corporate entities) in that national jurisdiction. All data controllers established in a country in question would need to sign up to the national HMB and pay a certain percentage of their profits (in case of for-profit entities) or their funding (in case of non-profit entities) to the HMB. HMBs would use these funds to cover their operating costs, and to establish a financial support mechanism from which people harmed by data use could make individual petitions to. HMBs would have a reporting branch and an investigative branch. The latter would also deal with petitions for financial and

[62] Similar types of funds can be seen when one looks to landlord risk mitigation funds – these are funds used to encourage landlords to rent to 'higher' risk tenants such as those with lower incomes or with previous evictions, and in some cases used to tackle issues of homelessness whereby if damages occur landlords are reimbursed up to a specified limit. See Katy Miller, 'Using Incentives to Engage Landlords: Risk Mitigation Funds' 15th April 2016 <https://www.usich.gov/news/using-incentives-to-engage-landlords-risk-mitigation-funds> accessed 8th April 2019.

other support provided for by the fund(s). The two branches would communicate and coordinate with each other, and indeed a central role of the individual petition system would be to provide information to the HMB about the nature and severity of experienced harms. This information could be used to assess patterns of harms caused by data uses and inform policy around good practice for big data.

To respond adequately to individual petitions, there would be multiple investigation panels within the HMB governed by a group of people who are independent from the data controller(s). This group – the steering committee – would consist of legal and data protection experts, but also include lay members, to ensure a varied membership and reduce the potential for regulatory or institutional capture. There would also be an appeal board within the HMB, where rejected petitions to the financial support mechanism could be appealed to and considered.

Individual Petition Procedure

Anyone who perceives that they experienced significant and undue harm by data use – either through the legal or illegal use of their own data, or somebody else's data[63] - and who wishes to report this, and/or who wishes to apply for financial support can do so via an informal individual petition submitted to the HMB. The petition will be assigned to an investigative panel within the HMB which will conduct a first review of the case to establish whether the case has the potential on the balance of probabilities to meet three requirements. These

[63] This is important also in the context of predictive analytics, where the analysis of other people's data can lead to the detection of patterns of undesirable characteristics that are then applied to a specific person; and if the person has this undesirable characteristic she can experience harm as a result. This mechanism is not new – it has existed in actuarial reasoning in insurance, criminal law, etc., for a long time. But the availability of wider sets of digital data covering more aspects of people's lives, and the rapid advance of computational tools and methods increase the scale of this problem. See: work on rights of secondary data subjects, including: Mark Taylor, Genetic data and the law: a critical perspective on privacy protections (Cambridge: Cambridge University Press, 2013).

requirements are that (a) the claimed harm is significant, that it is (b) undue and (c) that there is a plausible connection between the harm arising and the data use.

If these criteria have the potential to be met then the HMB would invite the applicant to provide further information, which the HMB then uses to make a decision on: (a) whether the case does indeed include *significant harm* to the applicant. Here we take significant harms to be those that would be considered significant to a reasonable person in the individual's position, or harms that a data controller could reasonably foresee that a particular data subject would consider to be significant. We adopt this hybrid subjective/objective test in recognition of the fact that the impact of a particular harm suffered is dependent on the personal characteristics and factors relating to the data subject. The assessment of (b) whether harm is *undue* depends on whether it can be justified by a legal requirement that can be considered fair. For example, if an individual experienced harm on the basis of being included in a criminal investigation on lawful grounds, and these grounds are fair in that they do not include an implicit bias against certain groups of people (e.g. penalising practices that are associated with poverty, for example), this harm would not be considered undue. Similarly, harm might not be undue if somebody signed up to a web service in full knowledge that the web service did not adhere to good data practice standards (if no laws were broken). While these are examples of considerations that would have bearing on the decision, whether or not a specific instance of harm is undue, including the fairness criterion, must be assessed on a case by case basis by the HMB. Discretion, we argue, is needed, because HMBs should be flexible to respond to harms stemming from big data practices which are still being uncovered, and which we cannot list in an exhaustive sense at the outset. Therefore, to maintain a responsive HMB system we cannot, *a priori*, prescribe a detailed framework for the assessment of harms. To mitigate against potential risks arising from such discretion, we envisage there being strong procedural transparency around how HMBs make decisions. This acts as a counterbalance to the lack of stringent *a priori* prescriptive rules

to determine the significance of harm, which will be assessed on a case-by-case basis. Once HMBs have been in place for several years it will then be possible to revisit the framework for assessing harms, and to distil greater guidelines around what should be considered a significant harm.

Finally, (c) whether it is *plausible* to assume a connection between the harm and specific instances of data use - this is a lower standard than would be employed within traditional legal remedies, and entails proving whether a reasonable person would view on the balance of probabilities that it was plausible that the harm could have resulted (but not proving that it necessarily did result) from the data use outlined.

When first established, we suggest HMB petitions would be confined to claims from natural persons. This is because HMBs are envisaged to mitigate harms for individuals and addressing power imbalances which arise between individuals and data controllers. Nonetheless, in saying this, we acknowledge that power imbalances may also negatively affect small and medium sized enterprises. Therefore, once established for natural persons and operating well in this context, at a later point the role of HMBs might be revisited to consider whether it would be feasible to expand to cover claims from such entities.

In terms of what types of harm(s) we expect people to report, this could be discrimination, stigma, or the loss of income following unauthorised re-identification.[64] But harm can also occur without undue re-identification of data, and even without any data having been taken from the person who was harmed. The practice of predictive analytics, for example, uses

[64] Re-identification is usually understood as the attempt to match anonymised or de-identified data with publicly available information/data to discover the individual to which the data belongs to. It has been shown that it is possible to re-identify even data that was previously believed to have been stripped of any identifying information. See Barbara Prainsack & Alena Buyx, 'A solidarity-based approach to the governance of research biobanks' (2013) 21(1) Medical Law Review 71-91.

insights from one group of people to discern patterns that will then be applied on other groups of people. As a result, those with 'risky' or otherwise undesirable characteristics can be excluded from certain offers or services, or they can become a target for heightened scrutiny or surveillance (e.g. patients who are identified to be particularly likely to overuse emergency rooms on the basis of predictive analytics).[65] Whenever such harms do not give grounds to a claim within data protection or tort law systems, because they are lawful, or because it is impossible for affected parties to prove what act or omission exactly caused the harm (e.g. due to data having been used by many different bodies and having been analysed in ways that are not open to scrutiny), individuals could be encouraged to approach a HMB.[66]

Alternatively, some people may wish to report harms they have experienced which they think are due to data use but may not wish to seek financial or other support, e.g. if the harm experienced is not significant but they still want it recorded and its causes addressed to prevent possible harms to others. A simple, informal feedback form would be set up for this purpose and would enable the HMB to fulfil an important monitoring role. This informal process could be used to prompt investigations for HMBs and if harms were found to be significant and repeated, findings could feed into regulatory frameworks to address these harms. To increase public knowledge of the potential for harms arising from data use, public awareness campaigns would need to be established by the government to educate the public, so they can identify possible harmful practices.

[65] For an overview of such risks from data use, see, Pam Dixon & Robert Gellman 'The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future' (2014) World Privacy Forum. <http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf> accessed 8th April 2019.

[66] It could be considered to also allow individuals to appeal to the HMB if they can make plausible case that pursuing legal remedies would be too onerous or costly to be reasonably possible.

Importantly, HMBs would thus *be subsidiary to* rather than replace existing legal protections, thereby filling the aforementioned regulatory gaps created by the digital era. Regulators and data controllers would still be required to minimise risks as far as reasonable and practical. Alongside this, HMBs would be established that offer financial and/or other support to those who are harmed by data use but for instance the claimed harms suffered would not meet thresholds for legal causation under existing legal remedies – for example, because no direct causal link between an action of a data controller and the experienced harm could be proven. In this way, HMBs fill the existing gaps in traditional legal systems, whilst also serving an important reporting function as patterns of harms caused by data use investigated can then be assessed and issues identified and fed back to regulators. HMBs thereby seek to complement available legal remedies in mitigating potential negative effects of data use, and also by providing feedback to data controllers on how systems and procedures could be improved. In EU member states, HMBs would complement, rather than compete with or duplicate, the role of the relevant Data Protection Authority (DPA), which is the independent public authority in each state that supervises the application of data protection law. While the main role of the DPA is to police data controllers, and while they are limited to cases that infringe existing laws, HMBs main role would be to support data subjects, and they would not be limited to instances of unlawful data use. In this area, HMB could assist in collecting evidence on the frequency, nature, and severity of harms occurring, and by analysing this information to help improve data protection, HMBs and DPAs could and should collaborate on this.

B. Financial Support Function

As noted, HMBs – particularly as an institution that people can appeal to also for financial support – do not aim to replace existing legal mechanisms for compensation or redress. Instead, they seek to *complement* legal systems by providing a low-threshold instrument for people who feel that the legal system did not, or cannot, address the harm that they suffered. HMBs could

make positive decisions on appeals even if data users did not infringe any laws or rules, and if HMBs gave financial support to applicants, these would not claim to provide full restitution or compensation for all losses resulting from the harm. In cases where the claimed harm could be quantified in financial terms, money paid out by HMBs would not claim to correspond with those figures either. Instead, money paid by HMBs would be understood as financial support, not necessarily corresponding with the extent of the actual harm. The amount provided would however aim to reflect the degree and type of harm suffered. In other words, the more significant the harm suffered, the higher the financial support offered. This would be assessed on a case by case basis. When the system is first set up, if concerns arose initially on the affordability of the system the financial supports offered could be set out as a percentage of losses/costs borne by the individual, e.g. 60% of actual loss/costs, which would reflect the fact that it is a supportive measure to the individual. It also reflects the informal nature of system which does not require legal costs, requires a minimal application process in terms of the data subject etc.

Liability under the HMB model is on a no-fault basis such that financial support is not dependent on proving a violation of law by the data controller. Instead it is based on proving a plausible connection between the actions (legal or illegal) of the data controller and the harm suffered. As noted, awards could be made from HMBs even if the actions of the data controller did not fall foul of any laws, which addresses concerns raised elsewhere on the need to have broader systems of accountability for data uses in the biomedical context.[67] In this way, HMBs are distinctly different from traditional constructions of liability e.g. in tort law where liability is generally premised on a breach of duty and standard of care, with a direct proof of causation between the act in question and harm arising. Instead, in the HMB context, if an individual

[67] Nuffield Council on Bioethics, 'The collecting. linking and use of biomedical research and health care: ethical issues' (April 2014), [4.46].

experienced harm, they could appeal to an HMB which would assess whether financial support is warranted based on the abovementioned criteria. In theory, claims could also be made in cases where data controllers' actions were not unlawful.

Three key legal questions arise under this framework, namely: (a) definition and construction of harm for the purposes of the HMB; (b) the subsidiary nature of HMBs and how the system is also designed to address harms falling outside existing protections, (c) in cases where the response of an HMB takes the form of support including e.g. financial payments, how the amount of financial support would be awarded. Taking each of these aspects in turn, the following can be said:

1. *Definition and construction of harm for purposes of HMB*

This section uses the European data protection framework as a case study to illustrate how the conception of 'harm' in HMBs differs from existing legal frameworks under the GDPR and the previous framework under the Data Protection Directive 95/46/EC. The precise role and influence of the GDPR over UK data protection laws after the UK leaves the EU is still uncertain, and will depend on the outcome of any Brexit deal.[68] Nonetheless, as of May 2018 the GDPR applies directly also in the UK until it leaves the EU likely later in 2019, and even after that transitional measures are still being negotiated. Furthermore, the UK's Data Protection Act 2018 came into effect in May 2018. This Act replaces the previous Data Protection Act 1998, applies GDPR standards and complements the GDPR by setting out

[68] This position is correct at time of writing 8th April 2019.If the UK leaves without a deal the GDPR will cease to operate in its current form, however, measures including the Draft Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019 have been drafted in the UK which require governmental approval, but which seek to further align UK data protection laws with the GDPR post-Brexit. This would allow the UK to apply to the European Commission for an adequacy decision on UK laws on whether they provide adequate protection for personal data in line with the GDPR. See discussion at Kingsley Napley, 'GDPR for the UK: Brexit and international transfers of personal data' (Lexology, 9th January 2019) https://www.lexology.com/library/detail.aspx?g=78e1c762-5c01-47d1-aa35-d2cff2e8991a

specific rules to supplement its application in the UK. Therefore, the domestic Act will ensure continued application of provisions contained therein which are related to the GDPR in the UK post-Brexit. Moreover, even if the GDPR does not apply directly within the UK post-Brexit, the territorial scope of the GDPR is broader than the previous Data Protection Directive 95/46/EC, and the GDPR applies to all data controllers/processers who are processing personal data of individuals resident in the EU, which is regardless of the controllers/processers place of establishment. Given the proximity of the UK to EU markets, this is still likely to apply to many UK data controllers/processers. This section briefly sets out the relevant provisions and remedies provided previously by the Data Protection Directive 1995 (by reference to how these were applied in the UK under the Data Protection Act 1998) considering how the GDPR improves upon these, and also noting the gaps remaining.

The UK's Data Protection Act (DPA) 1998, which brought the EU Data Protection Directive 1995 into national law, stipulated that harms from data use/mis-use were legally sanctionable under UK law if recognised under the relevant statute (DPA). Other legal actions were also relevant in this context, e.g. data use/mis-use may be subject to common law action for breach of confidentiality or tort of misuse of private information; or it may in some cases breach rights under the European Convention of Human Rights, of most relevance here is Art 8 (right to respect for private and family life).[69] Data controllers who did not abide by standards set out in the DPA 1998 were liable to sanctions. However, the threshold required to prove harm and the penalties imposed when harm was established left gaps for data subjects. For example, the

[69] Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014) http://nuffieldbioethics.org/wp-content/uploads/FINAL-Report-on-Harms-Arising-from-Use-of-Health-and-Biomedical-Data-30-JUNE-2014.pdf accessed 21 May, 2018, 28. See also discussion in: Article 29 Working Party, Opinion 4/2007 on the concept of personal data 01248/07/EN WP 136 available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf at 18.

penalties imposed were arguably not severe enough to effectively deter data misuse.[70] Moreover, and importantly in the context of big data, the provisions of the DPA 1998 did not apply to anonymous data, so no legal remedies were available if the use of anonymous data caused harm. The GDPR improves this situation by expanding the concept of personal data to include (at least some instances of) 'pseudonymised data', i.e. data that is neither anonymous nor identifying,[71] whereby one attribute in data (usually a unique identifier) is replaced with another as an extra security measure to reduce the risk that the data subject can be identified.[72] It thereby extends the potential for legal remedies for individuals who did not have access to them under the Data Protection Directive.[73] It remains to be seen, however, how this will be implemented in practice.[74] Also, the GDPR does not include fully anonymous data (i.e. data where it is assumed that no link to specific individuals can be made)[75] under the remit of personal data and thus leaves those who are harmed by the use of such data without redress. Furthermore, in distinguishing between anonymised and pseudonymised data a key question will be whether the risk of reidentification is reasonable,[76] which could provide a potential gap

[70] For a discussion, see (Out-Law.com, 'Jail sentence penalties for data breaches will be consulted on despite Government's scepticism' (11 October 2013) <https://www.out-law.com/en/articles/2013/October/jail-sentence-penalties-for-data-breaches-will-be-consulted-on-despite-governments-scepticism/> accessed 8th April 2019; Out-Law.com, 'Review of UK data protection: Should fines go over half a mil?' Out-Law 6 March 2014 <https://www.theregister.co.uk/2014/03/06/uk_review_of_data_protection_sanctions_threshold/> accessed 8th April 2019.

[71] Miranda Moubray et al, 'Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK' (2018) 34(2) Computer Law and Security Review 222-233.

[72] Article 29 Working Party, 'Opinion 05/2014 on Anonymisation Techniques' 10 April 2014, 0829/14/EN WP216 at 20

[73] Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014), 33.

[74] For instance, Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014), 34 who state at footnote 50 that "…there is concern over how the Regulation would impact (negatively) upon the processing of personal data for health and biomedical purposes. The Wellcome Trust has consistently opposed drafts of the GDPR, which purport to turn pseudonymous data into a subset of personal data that would interfere with publicly beneficial research from being carried out."

[75] Information that ''does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is no longer identifiable'.

[76] Gabe Maldoff, 'Top 10 operational impacts of the GDPR: Part 8 – Pseudonymization' (12th February 2016) https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/ accessed 8th April 2019.

in protection, depending on how 'reasonable' is interpreted in such contexts, and particularly given ongoing advances in technology. [77]

More generally, Graeme Laurie et al previously highlighted the narrow framing of harm under the DPA 1998 whereby individuals had to prove that harm experienced by them caused either financial damage or distress (emotional suffering) and that this was of a sufficient degree to constitute a breach of the DPA 1998. [78] Under the GDPR, recital 82 provides a right to compensation for any person who suffered "material or non-material damage as a result of an infringement of this Regulation", and compensation can be obtained from the data controller or the processor. There is no definition of non-material damage, however, recital 85 GDPR provides guidance, defining potential harms as:

> "physical, material or non-material damage to natural persons such as loss of control over their personal data or limitation of their rights, discrimination, identity theft or fraud, financial loss, unauthorised reversal of pseudonymisation, damage to reputation, loss of confidentiality of personal data protected by professional secrecy or any other significant economic or social disadvantage to the natural person concerned."[79]

It remains to be seen how such definitions of damage/harm will be applied under the GDPR in practice, but arguably, although broader than the DPD by the inclusion of non-material damage,

---

[77] There is a recognition that as technology develops data which has previously been annoymised, may become identifiable. The Irish Data Protection Commission states : "likely that more advanced data processing techniques than currently exist will be developed in the future that may diminish any current anonymisation techniques. It is also likely that more data sets will be released into the public domain, allowing for cross comparison between datasets. Both of these developments will make it more likely that individual records can be linked between datasets in spite of any anonymisation techniques employed, and ultimately that individuals can be identified." See https://www.dataprotection.ie/en/guidance-landing/anonymisation-and-pseudonymisation Reasonable measures must be taken against reidentification, but arguably this could not extend to pre-empting future specific technological advances.

[78] Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014), 3B and 3B3.

[79] See also recital 75; Art 29 Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679 ' Adopted on 3 October 2017 (Revised and Adopted on 6 February 2018) at p. 6

they could retain a relatively narrow construction in practice. Art 82(3) also provides an exemption for controllers/processers from liability if they can prove that they are "not in any way responsible for the event giving rise to the damage."

In contrast to this approach – and in addition to the aforementioned differences between HMBs and DPA acting under the remit of the GDPR– HMBs would seek to employ a broad definition and application of harm arising from data-use.[80] HMBs would recognise physical, psychological and financial harms experienced by individuals if on the balance of probabilities, it is *plausible* that this harm is connected to data use. Instead of looking towards a strict traditional legal classification of harm which generally relies on causality to the loss suffered, the HMB (a) would adopt a lower threshold to prove causality such that the individual need only prove there is a plausible link between the harm suffered and the data use(s). It would also not need to be a proven link between an action by a specific data controller and the harm. Instead, in recognition of the pervasive nature of data, it would be enough to prove a plausible link between multiple data uses and harm arising. (b) Drawing on Laurie et al's findings and recommendations,[81] HMBs would look at harm in the sense of how individuals were *'impacted'*

[80] For a discussion of cases on this, under the previous DPA and limits on see, Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014), 7B1F; See also: K.H. Jones, G Laurie, L Stevens, C Dobbs, DV Ford, N Lea, 'The other side of the coin: Harm due to the non-use of health-related data' (2017) 97 International Journal of Medical Informatics 43–51.

[81]Graeme Laurie et al, 'A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data', Report for Nuffield Council on Bioethics Working Party on Biological and Health Data and the Wellcome Trust's Expert Advisory Group on Data Access (June 2014), 161 who state: "To capture this, our soft evidence base conceptualised the notion of 'impact' arising from data use. Thus, for example, an individual might experience an impact if her/his data are used without permission, even if this is perfectly legal. Equally, organisations handling data might suffer an impact in trust and allegiance if individuals or groups whose data are held and used perceive an adverse impact through uses of which they disapprove. This is not to suggest that groundless concerns or abstract fears should drive information governance practices. Rather –as our soft evidence base suggests – the range of considerations about what might be construed as harmful is far wider than the law alone recognises. As such, the lesson is that due attention should be paid to possible impacts when using health and biomedical data, and to ensuring that governance mechanisms and actors within them have the ability to assess and, where appropriate, respond to data subjects' expectations."

by data use when assessing the significance of the harm looking at the extent to which this was harmful to that individual(s). This broader classification and application of harm is needed to support individuals in an era where decisions are increasingly supported – or even driven – by data. Also, whilst Art 82(1) refers to compensation being awarded based on "damage suffered" HMBs could go beyond this because the financial support that HMBs can offer are not limited to damage suffered but extend to other harms. Furthermore, although harms arising outside the GDPR, may be actionable in other ways, e.g. as breaches of human rights under the ECHR, HMBs provide a more expedient and less costly way for individuals to report and request support for such harms.[82]

Nonetheless, as noted above, the HMB system is complementary to existing legal protections under data protection frameworks and is expected to be used by individuals who do not have claims under traditional legal remedies. For instance, the HMB would take claims which fell outside the scope of current laws where e.g. harm resulted from anonymised data; where harm resulted to a secondary data subject, or where harm occurred from lawful data use. In such cases, affected individuals could bring a petition to the HMB. However, to avoid duplication, although individuals who have a traditional claim are not barred from applying to the HMB, individuals are not be able to appeal to HMBs if other legal actions are in progress e.g. human rights claims or claims under the GDPR. Instead, if a legal claim started in the course of a petition to the HMB, a stay on the petition would operate until the legal action was concluded. The HMB could then continue to consider the petition. However, as it is intended as a support mechanism for individuals and does not act as compensation, therefore if someone was compensated by the legal framework, the HMB petition would be unlikely to lead to further

[82] In this respect, the administrative costs of running the HMB would be monitored particularly at the initial stages of the system and would need to adapt accordingly. For example, if it transpired that individuals were more inclined to use this informal HMB facility rather than look to traditional legal remedies then there would, for instance, need to be a percentage increase in the amount that each data controller would pay to the HMB.

financial rewards as that individual's harm would already be financially mitigated against by law. Nonetheless, other supports might be granted to such individuals[83] and petitions would also be used to inform the governance feedback loop created by HMBs.

2. HMB Subsidiarity: Addressing harms falling outside existing protections.

As mentioned, despite provisions in the GDPR which seek to address some of the challenges posed by big data, important gaps remain, which HMBs could help address. For example, as noted above, Art 22 (1) GDPR states that:

> the data subject shall have the right not to be subject to a decision based *solely* on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Where profiling is defined in Art 4 GDPR as:

> any form of automated processing of personal data consisting of using those data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

However, the effect of the provision is limited by the fact that it refers only to processing of an individual's 'personal data' used in a manner to 'evaluate certain personal aspects' relating to

[83] Examples would include the recommendation by an HMB to reinstate credit worthiness that was negatively affected by association.

that person[84] and is only applicable where there is no human intervention evident in the processing. Only fully automated data analytics fall within this, and those not fully automated – even if they include humans in an relatively insignificant manner, e.g. to sign off the decisions of the machine – do not fall within the remit of the protection (for further exceptions see below).[85] As this article also relates to personal data as defined in the regulation it is unlikely - given the specific use of '*that* natural person' in the article - that this would apply to data of individuals used to make predictions which impact upon a secondary data user. Furthermore, the right does not apply if covered by the exceptions in Art 22(2) which allow automated processing if the decision is: necessary for the performance of a contract between data subject and controller; based on the data subject's explicit consent; or if it is 'authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests'.[86] Whilst the Art 29 Working Party guidance has highlighted limits placed on these caveats, such as having a narrow construction of necessity in the context of performance of a contract,[87] nonetheless, the practical effect of Art 22 is significantly curtailed by such caveats.

The GDPR also explicitly prohibits using sensitive data (or special categories of data) for automated decision-making purposes, unless the data controller has implemented measures to safeguard the data subject's rights,[88] and that the data subject has explicitly consented to

[84] For a discussion of provisions see O Proust, 'Getting to know the GDPR, Part 5: Your big data analytics and profiling activities may be seriously curtailed' FieldFisher (4 December, 2015) <http://privacylawblog.fieldfisher.com/2015/getting-to-know-the-gdpr-part-5-your-big-data-analytics-and-profiling-activities-may-be-seriously-curtailed/> accessed 8th April 2019.
[85] As noted above Art 29 Working Party guidance, endorsed by EDPB stated must be meaningful human action, not token. Art 29 Working Party, Guidelines on Automated Individual decision-making and profiling for the purposes of Regulation 2016/679 adopted on 3 February 2017 WP251.01 at 21.
[86] Art 22(2)(b).
[87] Art 29 Working Party guidance, endorsed by EDPB stated must be meaningful human action, not token. Art 29 Working Party, Guidelines on Automated Individual decision-making and profiling for the purposes of Regulation 2016/679 adopted on 3 February 2017 WP251.01 at 13.
[88] Sayers, Samantha & Drury-Smith, James (2016) "Legislative Comment: GDPR series: how to operationalise profiling for your organisation" 17(1) Privacy and Data Protection 3-6, 5.

processing, and the processing is for a legitimate aim.[89] Nonetheless, this fails to address the fact that even with such safeguards in place, harm can still occur, in view of the aforementioned effect of big data where people do not know what actions exactly led to harm, and where they sometimes do not even know that data were used that proved harmful to them.[90] Moreover, individuals may not have access to legal remedies, e.g. if it is a secondary data subject who is harmed.

The GDPR also recognises the risk of discriminatory decision-making arising from profiling,[91] These include Recital 75 which recognises that algorithms can be used in a way that causes indirect discriminatory effects for certain individuals 'even if those organisations had no knowledge of the discrimination and did not intend to discriminate"[92] and legal protections against discrimination could be employed in such contexts. Guidance from the UK Information Commissioner's Office emphasises the need for data controllers to take steps to prevent bias and discrimination resulting from profiling.[93] However, in this context, the focus is again on the processing of 'personal data'[94] and the scope and effect of this provision in the 'big data' era remains to be seen. Moreover, as is well known, despite the GDPR's goal to remove national differences in data protection standards, it leaves ample room for national derogations, and for different interpretations of key terms such as 'public interest' or 'appropriate safeguards'. In

[89] See Art 9(2)(a) and (g) as cited in Samantha Sayers and James Drury-Smith, 'Legislative Comment: GDPR series: how to operationalise profiling for your organisation' (2016) 17(1) Privacy and Data Protection 3-6, 5.
[90] Pam Dixon & Robert Gellman 'The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future' (2014) World Privacy Forum. <http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf> accessed 8th April 2019.
[91] See recitals 71 and 75 GDPR, see discussion in Ann Bevitt and Laura Dietschy, 'Legislative Comment: GDPR series: the risks with data profiling' (2016) 17(2) Privacy and Data Protection 7.
[92] Ann Bevitt and Laura Dietschy, 'Legislative Comment: GDPR series: the risks with data profiling' (2016) 17(2) Privacy and Data Protection 7.
[93]Information Commissioner's Office, 'Guide to the GDPR', available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/
[94] Defined in the regulation as: Art 4: 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;'

sum, the GDPR leaves plenty of room for individuals to be harmed in a big data context which HMBs could help address.

3. Financial Support Fund: Appeals Process

Turning next to the practical operation of the financial support function, three main questions arise which will be addressed briefly here.

i. Who could apply to the HMB for financial support and on what grounds?

Given the focus on collective risk management/benefits, any natural person could apply for financial support if they could make a plausible case that they have experienced significant and undue harm by data use. An important feature of HMBs is that it would also be open to appeals from secondary data subjects - individuals whose own data has not been used, but who have been harmed by the use of other people's data. Since there is no recognised relationship between the data controller and a secondary data subject under traditional legal approaches, it is questionable what duties the controller would have to such subjects within existing frameworks. However, third parties have been recognised as having rights in other contexts, such as within contract law,[95] or insurance law.[96] In the big data context, where 'secondary harms' will occur more frequently, effective protections are imperative in the recognition of individuals' commitment to big data, and to ensure fairness given the power disparity between individuals and data controllers. The openness of HMBs to providing support to secondary data subjects is an expression of collective responsibility for harm resulting from any kind of data use. Importantly, as noted, the HMB would initially only allow claims from natural persons and not from corporate entities.

[95] In the UK context, the relevant legislation is the: Contracts (Rights of Third Parties) Act 1999.
[96] In the UK context, the relevant legislation is the: Third Parties (Rights Against Insurers) Act 2010.

There would also be no time-limit on claims to the HMB. This is because general statutory limitation periods are designed for public policy reasons, namely: (1) it is unfair for companies or individuals to be challenged on basis of old allegations of wrongdoing as evidence to prove or refute a civil action may become difficult to verify/obtain if a substantial period has elapsed between the alleged wrongdoing and time of the action being raised; (2) there should be a certain period of time after which a wrongdoer should not have the threat of legal action.[97] However, HMBs are designed on the basis of no-fault and as noted applicants apply for support from the HMB not the data controller, and the requirement is based on plausible connection to the data use in the HMB context which will be lower requirement to establish than the chain of legal causation required for traditional actions. Moreover, HMBs would be an instantiation of a collective commitment to support those who were harmed by (legal or illegal) data use and an acknowledgement that not all risks are foreseeable in the context of big data. In this spirit, depending on the severity of the harm and reasons for not taking a claim earlier, financial support could still be made at a time removed from the initial data uses. However, there would need to be a justifiable reason for the delay in making a claim to the HMB in cases where the harm was discovered long before the claim was applied for.

ii. How would a claim be made to the HMB, and how would payments (if any) be assessed?

As briefly sketched above, a key feature of the HMB and its financial support mechanism is that it is simple to use and thus, bureaucracy would be kept as low as possible as a key objective. Therefore, a simple letter describing how a person has been harmed and how the person thinks this harm was caused (at least in part) by data collected or used by the data controller would be

[97] The relevant legislation is Limitations Act 1980 (as amended) applicable in England and Wales which provides for a three-year limitation period for claims for compensation for personal injury. For claims in tort other than for personal injury the limitation period is generally 6 years from the date of the damage is sustained.

sufficient to be considered by the HMB. The HMB would then conduct an investigation into the matter. This could mirror the approach adopted by bodies such as the Motor Insurance Bureau (MIB) compensation system which offers financial support to individuals involved in a motor accident where the other party was uninsured, or where the other party is untraceable (e.g. if they left the scene of the accident) or for UK residents involved in accidents with foreign registered vehicles either in the UK or EU.[98] Under the MIB scheme, claimants must submit a claim form[99] including details of the incident where they are claiming injury/loss arose from, and supporting documents/details including: witness/police reports etc. Following receipt of the form the MIB then conducts an investigation which includes: "establishing the facts; confirming the identity of those involved; obtaining independent reports from motor engineers or witnesses; obtaining a police report; contacting other bodies such as the DVLA, your insurer or a foreign bureau.'[100] Claims are dealt with within three months generally.

Similarly, under the HMB, an individual would submit a claim form to the national HMB, providing a description of the harm suffered and circumstances of this, alongside supporting evidence including for instance, medical reports etc. depending on the harm allegedly suffered. The HMB would then consider such supporting evidence and could request independent assessments/opinions from data experts in deciding on the claim, and/or enter a dialogue with the applicant to obtain further evidence or information. A key difference between the MIB scheme and the process envisaged for HMB is that under the former, compensation is fault based - it is only paid where fault is established on part of driver that the claimant considers responsible and if the claimant is wholly/partly responsible compensation may be reduced or not paid.[101] The HMB, in contrast, would work on a no-fault basis and financial support could

[98] See https://www.mib.org.uk/making-a-claim/what-we-do/ It remains to be seen how Brexit and leaving the EU will impact upon this.
[99] https://www.mib.org.uk/media/418096/mib_claim_form_v0618_v2a.pdfaccessed 8th April 2019.
[100] https://www.mib.org.uk/media/216242/your-guide-to-making-an-mib-claim.pdf accessed 8th April 2019, 11.
[101] Ibid, 11.

be paid even if the data use was outside the scope of relevant laws, provided sufficient harm arising from data use could be demonstrated, and that it was plausible this resulted from the data use.

If the HMB was satisfied that the harm suffered was significant, that it was undue, and that there was a plausible causation to data use, it could do any or all of the three following things: (a) acknowledge the harm and issue an apology to the applicant on behalf of the data controller(s); (b) feed information back to data controllers and policy makers, with the aim of improving procedures and rules to avoid such harm from occurring in the future; (c) make financial payments to support the harmed party under the financial support mechanism.

iii. How would the decision-making process operate?

HMBs would be governed by a steering committee independent from the data controllers that pay into the fund established by the HMB, which would develop a framework of criteria for decision-making by the HMB, in a transparent way and based on broad public participation. As noted above, HMBs would have appeals panels who hear and decide on appeals from residents of that country who complain of harm. Members of HMB appeals panels would include 'lay' members in their capacity as patients, etc., but also experts in data protection, marketing, security – depending on the size and remit of the HMB. Based on the decision-making framework, members of appeals panels would be free to consider any aspect that they deemed relevant provided the following conditions were met: (a) evidence of physical, psychological, financial, or reputational harm (b) a plausible case, on the balance of probabilities, would be made that the harm resulted, at least in part, from data collected or used by an organisation within the remit of the HMB. They would then decide what response would be adequate to the harm experienced by the appealing party, as noted above: an official acknowledgement of the harm with an explanation of what will be done to avoid that similar issues happening in the

future, and/or providing financial support. HMBs would need adequate financial resources to manage such a system and to ensure the independence of members to avoid issues of regulatory capture.

C. Evaluation and Feedback on Governance Systems

As noted, alongside the financial support mechanism, HMBs would have an advisory role and would be required to produce an annual report providing an overview of the types and distribution of claims to the fund that it investigated and the outcomes of these. This should be used to feed into good practice recommendations on data-use as the HMB would be in an ideal position to assess how harms materialised in the previous year by having an overview of such challenges. In this role as a collector and analyst on the evidence on the frequency, nature, and severity of harms resulting from data use, HMBs could collaborate with DPAs – although also here, their role would go beyond the remit of the DPA which is limited to harms arising from unlawful data use.

The feedback role of the HMB would form part of a process of reflexive governance.[102] If, for instance, multiple claims were received in relation to a specific harm being caused by a data-practice, this should highlight a pattern which would be identifiable by the HMB. The HMB could further investigate and highlight harm causing practices and could also seek to develop strategies of improvement and/or mitigation.

The feedback role is an important reason for having HMBs at a national level, since this would allow patterns of use to be identified across the country which should lead to deeper insights than for instance if HMBs were to operate on a specific industry or organisational level.

[102] Graeme Laurie, 'Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law' (2011) 130(3) Human Genetics 347-56.

Moreover, as harms are increasingly being triggered by issues that arise when datasets are shared across different areas and the combination of multiple data-sets, having such an overarching national body would allow a more coherent and comprehensive picture to be drawn. This should also help avoid the risk of some problems being missed, since a too narrow focus on a particular sector/industry might lead to ignoring the pervasive nature of the data use and harms triggered in the context of big data.

In addition, having a national body would allow comparisons to be made in terms of the types of petitions which are refused by the HMB for support. If a pattern of refusals occurred in a particular context over time, it might require the HMB to reconsider how such petitions are being evaluated, or what should be fed back to corporate data users in such contexts to ensure that a gap between what individuals feel is acceptable and data uses occurring would be recognised. Where petitions were refused because harm was not plausibly connected to the data use, but a similar pattern was complained of by multiple petitioners, this could illustrate misplaced fears individuals might have about data use – which, if allowed to continue, could hamper beneficial progress of big data. If such patterns were detected by HMBs, they could be addressed by reassuring individual petitioners that harm was not plausibly connected with data use and also by further educational campaigns demonstrating how big data operates.

Art 35 GDPR provides that for processing using new technologies likely to result in high risks to rights of natural persons, controllers are required to conduct assessments on the protection of personal data.[103] However, gaps will remain, given the difficulties in fore-sighting

[103] See Information Commissioner's Office guidance on Data Protection Assessments at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ ; Art 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, 17/EN WP 248 rev.01 as adopted by EDPB available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

technologies and the challenge specific to big data. The feedback role intended by the HMB fills this gap and is intended to complement the forward-looking role of such impact assessments. This feedback role would also complement the role of Supervisory Authorities in each country under Art 57 GDPR. Such authorities are responsible for monitoring and enforcing the GDPR, and also "monitor relevant developments, insofar as they have an impact on the protection of personal data, in particular the development of information and communication technologies and commercial practices."[104] HMBs would complement this role, as while the Supervisory Authorities are likely to focus on issues arising due to failure to comply with the GDPR, HMBs look beyond this by considering harms arising regardless of the lawfulness of processing under the GDPR, and for all data uses not just uses of personal data.

D. Operational Overview

As noted, HMB could be established at the national level. A key feature is that they would be independent of the organisation, that is, the data controller, using the data and who review appeals from people who claim they have been harmed by data use. HMBs would be funded by a set percentage of the budget of each research project or of the overall institutional budget (depending on the form of organisation in question) being set aside for this, initially set, e.g., at 1%.

We also recognise that part of the issue with the governance of big data is that the sharing and integration of datasets between institutions and organisations can give rise to harms that may occur from the combined use of datasets from different people across borders. In recognition of this, and of that fact that data given to one organisation (provided appropriate consent was given) can be used for other purposes which could create knock-on benefits in other contexts

[104] Art. 57(1)(i) GDPR.

and jurisdictions, we suggest that if data users reside in several countries, the individual would apply to the one where she herself is a resident.[105] (While in this article we have used UK/European laws as an exemplar, we would hope that gradually other countries would adopt similar harm mitigation institutions and instruments.)

We envisage that the operation of HMB would be an independent arm of, and overseen by, the Information Commissioner's Office in the UK and equivalent bodies in other jurisdictions. Furthermore, there would be a national advisory committee/board dealing with HMBs which would have annual meetings and issue annual reports to the public. Representatives of data controllers would be invited to attend such meetings, and this would provide a forum for delegates to meet and exchange experiences or address common issues arising. Overtime, sub-committees might also be established to provide a more specific forum tailored to particular industries, e.g. the health context. However, given that many harms are systemic, these committees would not replace the national meetings and instead would be designed more as awareness raising meetings translating the overarching issues into relevant contexts for each industry. The advisory board would receive annual reports from the HMB steering committee and appeals panels and use these to develop national standards/guidance documents. Over time, HMBs at a European or even international level could also be established, recognising that much research is not nation specific, and data flows across jurisdictions.

CONCLUSION

Novel practices of curating, storing, and using digital data require new ways of thinking about data governance. The individual focus in legal frameworks is no longer sufficient to capture the

[105] This model rests on the assumption that on balance, the proportion of petitioners claiming harms for which financial support would be paid out corresponds with the proportion of funds coming into national HMFs via corporations residing in these countries (for example, Finnish users would experience harms from data use by a company residing in the USA). When this is not the case, an international mechanism to balance costs between countries would need to be implemented. Within Europe this could be addressed by EU law.

interests at stake or tackle the power asymmetries and inequities in costs and benefits of data use in the digital era. Moreover, key requirements of traditional legal remedies for corporate misuse of data – such as proving fault or causality – are no longer feasible in an era where data use is regularly not traceable. In addition, significant harms regularly occur from lawful data use. In an effort to increase collective responsibility for harms that people experience from data use, whether lawful or not, we have sketched out the instruments of HMBs. We argue that HMBs provide a mechanism to address some of the power asymmetries that mark data use in the digital era. They provide support for individuals harmed by data-use/mis-use by offering mechanisms of financial support to individuals where no existing legal remedies are available. They also provide a mechanism for the reporting of data governance issues, thereby offering a reflexive governance tool which is vital for emerging areas of data governance and particularly for big data in the digital context given the pace at which this area is developing. We expect there to be aspects of HMBs that need further elaboration and refinement before these bodies can be considered for implementation. Towards this end, we hope this article will stimulate debate and inspire colleagues to help us improve and develop this idea further.